DOCUMENT RESUME

ED 466 777 TM 034 298

AUTHOR Norris, Deborah
TITLE Calculating Confidence Intervals for Effect Sizes Using Noncentral Distributions.
PUB DATE 2002-02-16
NOTE 20p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Austin, TX, February 14-16, 2002).
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Software; *Effect Size; *Statistical Distributions
IDENTIFIERS *Confidence Intervals (Statistics)

ABSTRACT

This paper provides a brief review of the concepts of confidence intervals, effect sizes, and central and noncentral distributions. The use of confidence intervals around effect sizes is discussed. A demonstration of the Exploratory Software for Confidence Intervals (G. Cuming and S. Finch, 2001; ESCI) is given to illustrate effect size confidence intervals for the single and two-sample cases. (Contains 3 figures and 11 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 466 777

Running head: CALCULATING CONFIDENCE INTERVALS FOR EFFECT SIZES

Calculating Confidence Intervals for Effect Sizes Using Noncentral Distributions

Deborah Norris

University of North Texas

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

D. Norris

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM034288

Paper presented at the annual meeting of the Southwest Educational Research Association, February 16, 2002, Austin, TX

BEST COPY AVAILABLE

Abstract

This paper provides a basic review of the concepts of confidence intervals, effect sizes, central and noncentral distributions. Specifically, the use of confidence intervals around effect sizes is discussed. A demonstration of the *Exploratory Software for Confidence Intervals* (ESCI) is given to illustrate effect size confidence intervals for the single and two-sample case.

Calculating Confidence Intervals for Effect Sizes Using Noncentral Distributions

There is much debate regarding the utility of statistical significance testing as a means of testing research effects (cf. Harlow, Mulaik, & Steiger, 1997; Henson & Smith, 2000; Vacha-Haase, Nilsson, Reetz, Lance, & Thompson, 2000). Recommendations have been made to use other statistical methods to evaluate data, such as the reporting and interpretation of effect sizes. Such practice is critical to good statistical methodology. As the American Psychological Association (APA) Task Force on Statistical Inference noted (Wilkinson & APA Task Force on Statistical Inference, 1999),

> It is hard to imagine a situation in which a dichotomous accept-reject decision is better than reporting and actual *p* value or, better still, a confidence interval. Never use the unfortunate expression "accept the null hypothesis." Always provide some effect-size estimate when reporting a *p* value (p.599).

The Task Force went on to state, "*Always* present effect sizes for primary outcomes" (p.599, emphasis added).

Influenced by the Task Force report, the fifth edition of the *APA Publication Manual* (APA, 2001) called the "failure to report effect sizes" a "defect in the design and reporting of

research" (p.5). Regarding effect size and strength of relationship the Manual stated,

> For the reader to fully understand the importance of your findings, it is almost always necessary to include some index of effect size or strength of relationship in your Results section. You can estimate the magnitude of the effect or the strength of the relationship with a number of common effect estimates… The general principle to be followed…is to provide the reader not only with information about statistical significance but also with enough information to assess the magnitude of the observed effect or relationship. (pp.25-26)

The fifth edition *Manual* (APA, 2001) also made comment on the role of confidence intervals (CIs) in result interpretation:

> The reporting of confidence intervals (for estimates of parameters, for functions of parameters such as differences in means, and for effect sizes) can be an extremely effective way of reporting results. Because confidence intervals combine information on location and precision and can often be directly used to infer significance levels, they are, in general, the best reporting strategy. The use of confidence intervals is therefore *strongly recommended.* (p.22, emphasis added)

The recommendation of the APA Task Force for the interpretation of both (a) effect sizes and (b) confidence intervals leads quite naturally to a conclusion that (c) confidence intervals about effect sizes would be quite useful (Thompson, 2001). The *APA Manual* (APA, 2001) also alludes to the possibility of CIs around effect sizes.

A special section of *Educational and Psychological Measurement* (Vol. 61, No. 4) is devoted to CIs around effect sizes. In this issue, Cumming and Finch (2001) presented new software that illustrates the value of CIs around effects for single and multiple studies. The software runs under Microsoft Excel, is user friendly, is quite reasonably priced, and is called *Exploratory Software for Confidence Intervals* (ESCI, pronounced "esky").

This paper will show how CIs can be used around effect sizes, focusing on Cohen's *d*. However, unlike typical CIs around many other statistics such as the mean, CIs for standardized effects require the use of noncentral *t* distributions.

Effect sizes

Effect sizes refer to indices used to indicate the magnitude of an obtained result or relationship (Fraenkel & Wallen, 1996). Two broad categories encompass effect size indices. The two broad categories are characterized by (a) directly examining differences between means (e.g., Cohen's *d*),

BEST COPY AVAILABLE

or (b) how much of the variability in the dependent variable is accounted for by variation in the independent variable(s). [i]Examples of variance-accounted-for effects include eta squared, omega squared, multiple R squared, and adjusted multiple R squared.

Cohen's *d*

Cohen's *d* is a mean or mean difference, standardized via division by the pooled standard deviation (Cumming & Finch, 2001).

$$\text{Cohen's } d = \frac{\mu_1 - \mu_2}{\sigma}$$

The interpretation of mean differences using Cohen's *d* is in terms of standard deviation units. According to Cohen (1988), *d* values of 0.2, 0.5, and 0.8 can be roughly regarded as small, medium, and large effects, although researchers too often rigidly employ these rules of thumb.

For a single sample case the following formula is applicable to generate *d* from a *t*-value:

$$d = \frac{t}{\sqrt{n}} \quad .$$

The utility of Cohen's *d* is expanded in relation to Pearson's *r* (point-biserial) by the following formula (given the two populations are of equal size).

$$r = \frac{d}{\sqrt{d^2 + 4}}$$

Because *d* is a ratio of two quantities (mean difference and standard deviation) noncentral *t* distributions are appropriate to compute accurate confidence intervals about standardized effect sizes (Cumming & Finch, 2001).

Central and Noncentral *t* Distributions

Prior to discussing CIs about standardized effects it is necessary to address central and noncentral *t* distributions. The familiar *t* distribution is really a special case of a broader class of distributions called "noncentral" distributions (Thompson, 2001). Typical inferential techniques are based on central distributions like the *t* distribution (Cumming & Finch, 2001). Central *t* distributions are a family of distributions based on degrees of freedom. The central *t* distribution is symmetrical, centered at zero, and approaches the standard normal distribution as sample size increases (Thompson, 2001). Central *t* distributions arise when a normally distributed variable with a mean of zero is divided by an independent variable closely related to the χ^2 distribution (Cumming & Finch, 2001). In contrast, noncentral *t* distributions arise when a normally distributed variable with a mean not equal to zero is

divided by an independent variable closely related to the χ^2 distribution.

Noncentral *t* distributions are not necessarily symmetrical and the degree of skewness depends on the noncentrality parameter (Δ). The noncentrality parameter is the distance by which the mean of the normal distribution is displaced from zero and can be estimated by multiplying *d* by the square root of *n* (Cumming & Finch, 2001). The noncentral *t* distribution is centered at approximately Δ, especially if Δ is small and the degrees of freedom is large (Cumming & Finch, 2001). When Δ is equal to zero, then the noncentral *t* distribution is exactly the same as the familiar central *t* distribution (Thompson, 2001).

Properties of the noncentral *t* distributions are illustrated using the ESCI software (workbook **Noncentralt**). Figure 1 presents a comparison of central *t* distributions (pictured on the left) with noncentral *t* distributions (pictured on the right.

INSERT FIGURE 1 ABOUT HERE

The ESCI program allows for manipulation of degrees of freedom and the noncentrality parameter (Δ) to show changes in the shape of the distribution. The following summary, based heavily on

Cumming and Finch (2001), identifies some of the properties of noncentral *t* distributions that are illustrated with the ESCI program:

- When the noncentrality parameter is equal to zero, noncentral *t* distributions are exactly the same as central *t* distributions, which means symmetric with a mean of zero.
- The noncentral *t* distribution is centered at approximately the value of the noncentrality parameter except for small degrees of freedom.
- The shape of noncentral *t* distributions will always be skewed unless $\Delta=0$.
- For a particular noncentrality parameter, as the degrees of freedom increase the shape of the curve is less skewed and more closely resembles central *t* distributions in shape.
- For a particular noncentrality parameter, as the degrees of freedom decrease the shape of the curve becomes flatter and more skewed.
- As the absolute value of the noncentrality parameter increases so does the extent of skewness associated with the distribution.
- The noncentrality parameter can be positive or negative with the outward tail larger with larger noncentrality parameters.

Confidence Intervals

Fidler and Thompson (2001) stated that more accurate definitions frame a CI as one interval from an infinite or at least large sample of CIs for a given parameter in which 1-α% (often 95%) of the intervals would capture the population parameter. CIs give a best point estimate of the population parameter of interest and an interval about that to reflect likely error-the precision of the estimate (Cumming & Finch, 2001). It is not certain if the confidence interval includes the true value of the parameter of interest unless the confidence level equals 100%. The width of the CI depends on the confidence, or probability, level. All else constant, large probability levels result in wider CIs. The width of the CI also speaks to the precision of the results. Smaller CIs usually infer more precision and less error associated with the results and therefore more confidence with the results.

CIs for Cohen's *d*

Constructing CIs for Cohen's *d* is not straightforward and requires the use of noncentral distributions. This is because estimating the CI involves a ratio of two parameters, mean (μ) and standard deviation (σ), which changes as either μ or σ change, and not simply for a single estimate such as μ using a

BEST COPY AVAILABLE

fixed estimated σ. This cannot be done by simply pivoting a probability statement (Cumming & Finch, 2001).

Calculating the CI for the standardized effect size *d* first involves constructing CI for the noncentrality parameter. Since there is no formula that can directly give the upper and lower bounds for the CI of the noncentrality parameter it is necessary to use computer software that uses an iterative algorithmic search to estimate these boundaries (Cumming & Finch, 2001). Once the upper and lower bounds for the noncentrality parameter are identified the CI for *d* can be calculated by dividing the upper and lower bounds by the square root of *n*.

ESCI for One and Two Group Designs

ESCI (**CIdelta**) is an interactive program that allows the user to calculate and display CIs for standardized effect sizes. Figure 2 illustrates a confidence interval for Cohen's *d* for a single group (i.e., single sample *t*-test). Hypothetical data (n = 30) were entered into the left side of the spreadsheet and the program calculated the mean and standard deviation as seen near the top of the Figure. The observed sample distribution is given at the top of the plot area and a scale of possible Cohen's *d* values is presented at the bottom of the plot area. With a "nil" null hypothesis of μ = 0 (which can be changed in

the program), the observed Cohen's *d* was 5.013, with $t(29) = 27.46$ and a *p* value practically zero.

Using the iterative process based on noncentral *t*, the CI is generated around the observed Cohen's *d*. The CI bar is seen in the plot area of Figure 2 and give a range of 3.677 to 6.341 (with a 95% confidence level). As noted, this bar would become wider if we were to use a 99% confidence level. This graphic informs us not only regarding the sample estimate of *d*, but also our level of precision in estimating it. The amount of useful interpretive information here is much greater than a single *p* value or even just the *d* effect size.

INSERT FIGURE 2 ABOUT HERE

Figure 3 presents results for a two sample design (i.e., two sample *t*-test). Again, hypothetical data were entered in the left side of the spreadsheet for Group 1 and 2 ($n_1 = n_2 = 30$). Means, standard deviations, and CIs for the means are then calculated. The mean difference and the CI for that difference are also presented. In the plot area, one can visualize the distributions, and the 95% CIs (this level can be changed) for the two means. The scale for the mean differences is provided and illustrates the possibility of both positive and negative

mean differences. Finally, the *d* scale is at the bottom of the plot area and the CI for the observed *d* of -0.501 is given as -1.013 to 0.0149. Of course, with a "nill" null of a zero mean difference, this effect is not statistically significant at the 95% confidence level, t(58) = -1.942, p = 0.057.

INSERT FIGURE 3 ABOUT HERE

The following conclusions, based heavily on Cumming and Finch (2001), are illustrated in the one and two case designs as depicted in Figures 2 and 3:

- The upper scale on the plot area shows the CI for the population mean μ.
- The CI for the population mean is symmetric and centered on the observed mean.
- The lower scale on the plot area shows the CI for the population standardized effect size δ.
- The CI for the population standardized effect size δ is not symmetric about the mean *d* due to the need to use noncentral *t* distributions.
- The standardized effect size measures a standardized distance from a particular chosen reference μ_0. If μ_0 is changed, there will be no change for CI about the

BEST COPY AVAILABLE

population mean but there may be a noticeable change for the CI about δ.

- The *d* scale depends on the placing of the μ_0, which sets the zero for the scale. It also depends on the standard deviation for a particular sample.
- The standard deviation for a particular sample sets the unit on the lower *d* scale.
- If different independent samples (with different data) were identified and compared to the illustrated one and two group designs, it would be expected that the CIs would be different, the standard deviation for the sample would be different, which would affect the units on the *d* scale which in turn would make the CI for δ different.

Summary

This paper provided an introduction to CIs about standardized effect sizes. A number of basic concepts including the use of noncentral *t* distributions were reviewed. The ESCI software program was used to simulate and illustrate CIs about single and two independent group designs. CIs around effect sizes (*d* and otherwise) can be extremely useful in both (a) result interpretation and (b) meta-analytic thinking that places our single point estimate in context regarding the population parameter.

References

American Psychological Association. (2001). *Publication Manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Cohen, J. (1988). *Statistical power analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cumming, G. (2001). Exploratory Software for Confidence Intervals [Computer software]. Bundoora, Australia: La Trobe University. Retrieved from http://www.psy.latrobe.edu.au/esci

Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement, 61*, 532-574.

Fraenkel, J., & Wallen, N. (1996). *How to Design and Evaluate Research in Education* (3rd ed.). New York: McGraw-Hill.

Fidler, F., & Thompson, B. (2001). Computing correct confidence intervals for ANOVA fixed- and random-effects effect sizes. *Educational and Psychological Measurement, 61*, 575-601.

Harlow, L.L., Mulaik, S.A., & Steiger, J.H. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.

Henson, R., & Smith, A. (2000). State of the art in statistical significance and effect size reporting: A review of the APA Task Force report and current trends, *Journal of Research and Development in Education, 33*, 285-296.

Smithson, M. (2001). Correct confidence intervals for various regression effect sizes and parameters: The importance of noncentral distributions in computing intervals. *Educational and Psychological Measurement, 61*, 605-632.

Thompson, B. (2001, April). *Using the descriptive bootstrap to investigate effect replicability (with a comment about confidence intervals for effect sizes)*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Vacha-Haase, T., Nilsson, J.E., Reetz, D.R., Lance, T.S. & Thompson, B. (2000). Reporting practices and APA editorial policies regarding statistical significance and effect size. *Theory & Psychology, 10*, 413-425.

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604. [Reprint available through the APA Home Page: http://www.apa.org/journals/smp/amp548594.html]

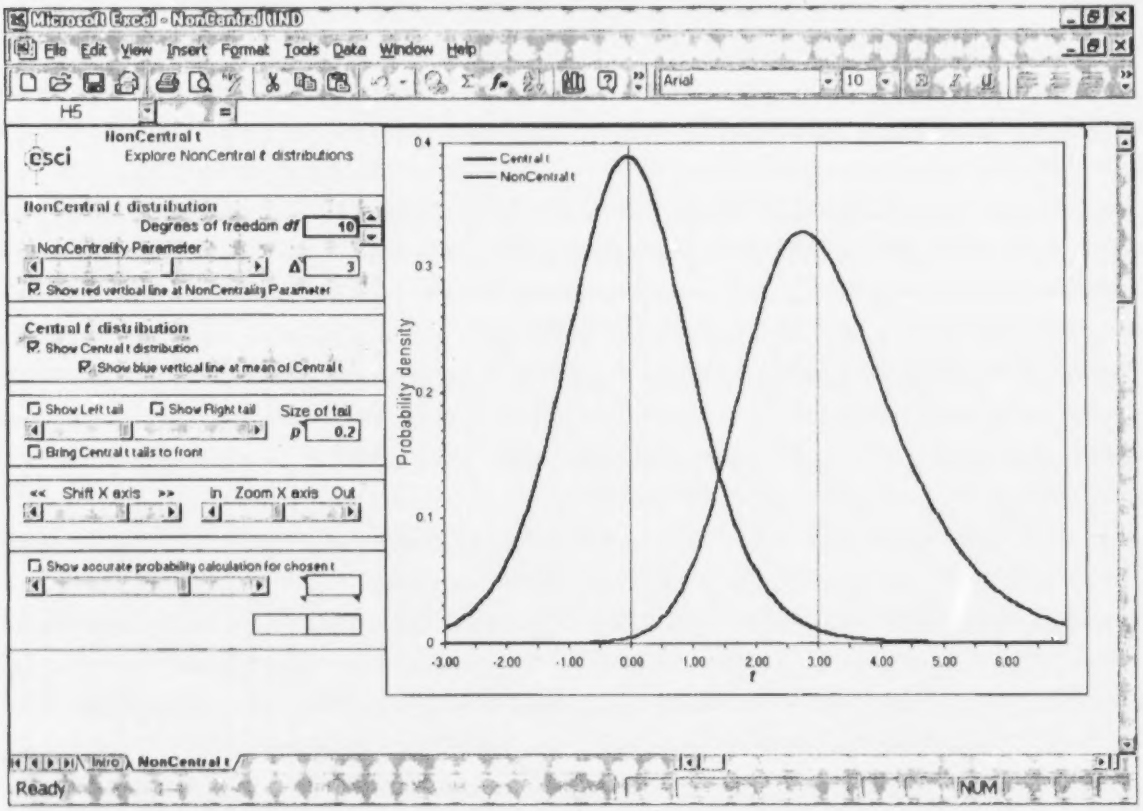

Figure 1. Central and noncentral *t* distributions.

BEST COPY AVAILABLE

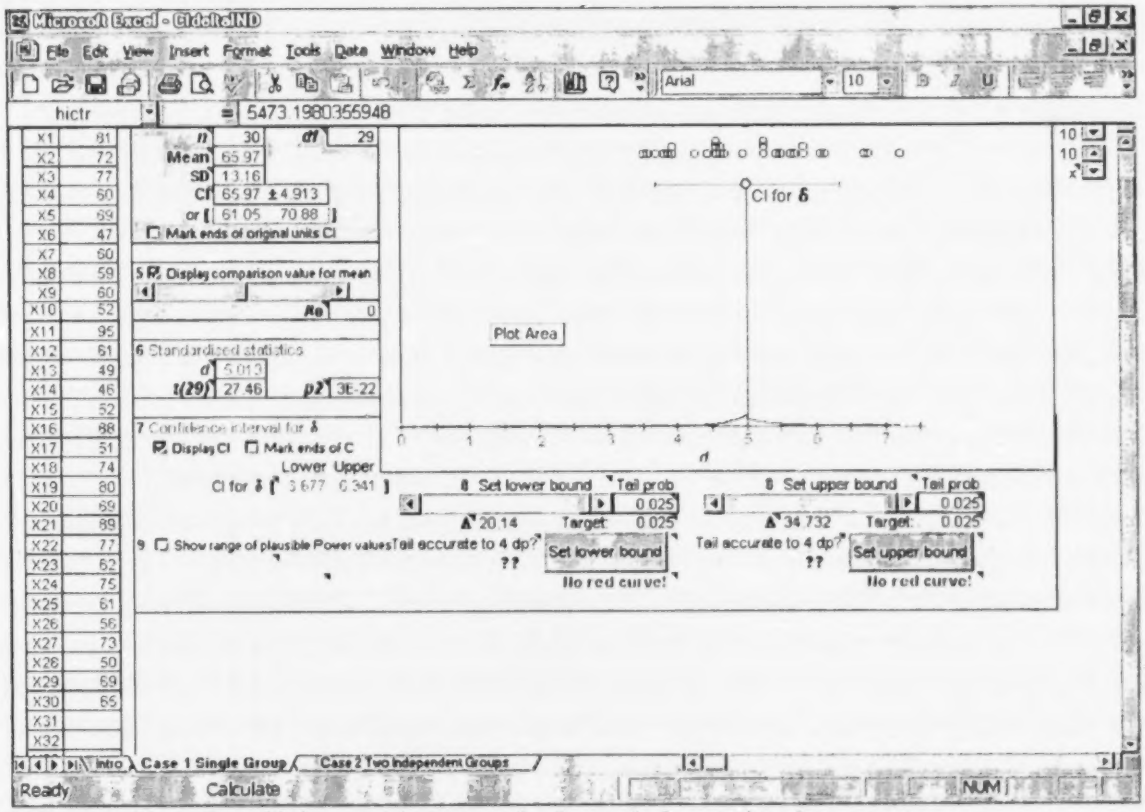

Figure 2. Confidence interval for *d* for single group.

BEST COPY AVAILABLE

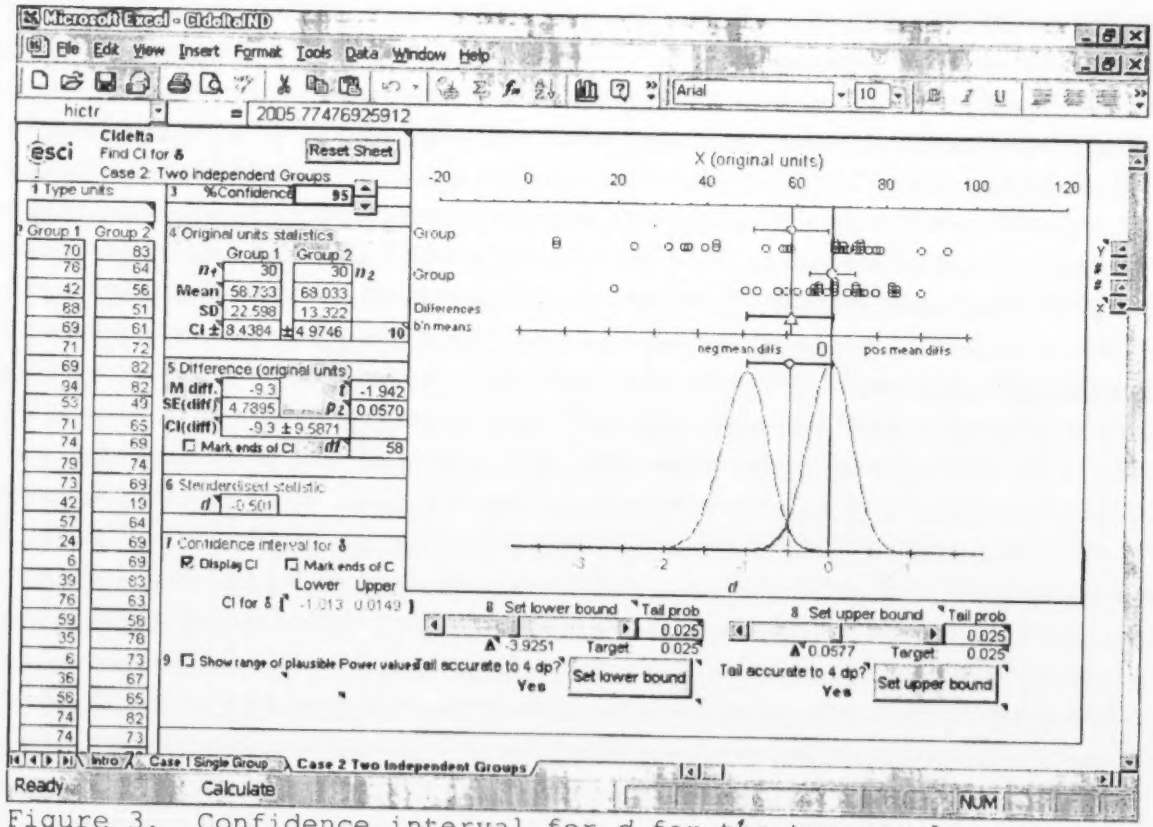

Figure 3. Confidence interval for *d* for the two sample case.

BEST COPY AVAILABLE

TM034288

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Calculating Confidence Intervals for Effect Sizes Using Noncentral Distributions	
Author(s): Deborah Norris	
Corporate Source: University of North Texas	Publication Date: Feb. 2002

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☒

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Deborah Norris	Printed Name/Position/Title: Deborah Norris / Res. Associate
Organization/Address: 1073 Springwood Dr. Saginaw, TX 76179	Telephone: 817 306 1715 FAX:
	E-Mail Address: deb3666@aol.com Date: 2/16/02

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 9/97)
PREVIOUS VERSIONS OF THIS FORM ARE OBSOLETE